

Astroinformática (AS4501) - Tarea 1

September 5, 2017

Fecha de entrega: Viernes 15 de Septiembre

Formato: Jupyter Notebook (.ipynb)

1. Distribuciones (3 pt)

- Construya un emisor virtual de fotones, que emita un fotón cada X milisegundos, donde X es un variable aleatoria con distribución Laplaciana. Utilice una escala suficientemente grande para evitar problemas de discretización de milisegundos. Luego, hacer un conteo de estos fotones (N) en un intervalo de tiempo T . Verificar que, para experimentos independientes, N sigue una distribución Poissoniana. (1 pt)
- Un método para comparar dos poblaciones basado en muestras independientes de ellas es la prueba de t-student. Para ello se utiliza el estadístico t^* :

$$t^* = \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

donde s_p es un estimador de varianza común:

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}},$$

\bar{x}_1 y \bar{x}_2 son los promedios de la muestra, n_1 y n_2 son el tamaño de las muestras y s_1 y s_2 son las desviaciones estándar de las muestras. El estadístico t^* evaluado en la distribución acumulativa de t-student da el grado de confianza del rechazo de la hipótesis nula que ambas muestras provienen de la distribución con mismos promedios. Es decir para un nivel de significancia del rechazo del 5% a que $\bar{x}_1 - \bar{x}_2 < 0$, t^* debe ser menor a -1.734.

Considerando estrellas de la secuencia principal a menos de 30 parsecs que sean anfitrionas de planetas según exoplanet.eu, responda:

- estrellas tipo K tienen en promedio planetas menos masivos que estrellas tipo G a un nivel de significancia del 5%?
- estrellas tipo K tienen en promedio planetas mas cercanos que estrellas tipo G a un nivel de significancia del 5%?

Notar que la prueba de t-student asume que las poblaciones tienen una distribución Gaussiana. Graficar la distribución de masa y de eje semimayor para percartarse que este requerimiento es mucho más certero al considerar el logaritmo de ambos parámetros.

(2 pt)

2. Análisis espacial (3 pt)

Una forma de estudiar la distribución espacial de galaxias es la función de correlación entre dos puntos (*two-point correlation function*). En esta pregunta se le pide construir la función de correlación entre dos puntos de la distribución de galaxias en la simulación *Millenium simulation*.

Los datos pueden ser bajados desde esta página (filtros SDSS):

<https://wwwmpa.mpa-garching.mpg.de/galform/agnpaper/>

- Para todas las preguntas siguientes comience con una región suficientemente pequeña y aumente su tamaño gradualmente, e.g. un cubo de lado creciente, monitoreando el uso de memoria.
- Construya histogramas para las magnitudes en i y colores $r - i$
- Divida la muestra en tres rangos de magnitudes representativos y tres rangos de colores representativos
- Construya y grafique la función de correlación entre dos puntos para cada una de estas divisiones usando al menos dos métodos diferentes.
- Para diferentes tamaños de muestras estime el tiempo de cálculo y grafique el tiempo de cálculo en función del tamaño de la muestra (es posible que no pueda calcular la función con toda la muestra)

¿Cómo realizaría el cálculo con la muestra completa?

Nota: para más información ver <http://www.nature.com/nature/journal/v440/n7088/full/nature04805.html>